DOCUMENT RESUME

ED 230 541 SP 022 383

AUTHOR Roth, Rod; Hanner, Austin
TITLE Assessing Minimal Competence for Certification with the National Teacher Examinations.
PUB DATE 20 Nov 82
NOTE 18p.; Paper presented at the Annual National Inservice Conference of the National Council of States on Inservice Education (7th, Atlanta, GA, November 19-23, 1982).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Descriptive (141)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Educational Legislation; Higher Education; Job Analysis; *Minimum Competency Testing; *National Competency Tests; Program Validation; Racial Discrimination; Relevance (Education); Standardized Tests; State Standards; Teacher Certification; Teacher Education Curriculum; Teacher Education Programs; *Test Interpretation; *Test Theory; Test Validity
IDENTIFIERS *National Teacher Examinations

ABSTRACT

This paper presents the legal rationale for using the National Teacher's Examination (NTE) validation procedure, which was used in Arkansas to determine cut-off scores for teacher certification. A review is given of the decision, by the United States Supreme Court, upholding use of the NTE by South Carolina in that state's teacher certification system. The points are made that it was not the intent of the testing system to discriminate and that graduation from an approved teacher training program was not a viable alternative to the use of the NTE for certification purposes. A review is also presented of Educational Testing Service (ETS) validation procedures that determine the validity of test questions as they relate to job requirements rather than to the quality of a teacher training program. The point is made that Uniform Guidelines on Employment Selection Procedures require identification of competencies important to a successful performance in the job for which candidates are to be evaluated. Actual directions given to educational experts who served as judges for the Arkansas NTE study of area examinations are presented. Also included in this paper are descriptions of data analysis procedures used in the Arkansas study. (JD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED230541

ASSESSING MINIMAL COMPETENCE FOR CERTIFICATION WITH THF NATIONAL TEACHER EXAMINATIONS

Dr. Rod Roth, Dean
College of Education
The University of Alabama

Mr. Austin Hanner, Coordinator
Teacher Education and Certification
Arkansas Department of Education

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Rod Roth
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

A paper presented at the Seventh Annual National Inservice Conference of the National Council of States on Inservice Education in Atlanta, Georgia on November 20, 1982

SP 022 283

Introduction

This paper presents the legal rationale for using the National Teacher's Examination (NTE) validation procedure that was used in Arkansas to determine cut-scores for Arkansas certification. The paper also presents the actual directions given to certified practitioners from Arkansas public schools and faculty members from colleges and universities in Arkansas who served as judges for the Arkansas NTE study of area examinations. The paper concludes with the data analysis procedures used in the Arkansas study.

Legal Rationale

This section of the paper presents the legal rationale for using the validation procedure that was used in Arkansas. In order to build the rationale, the NTE decision in South Carolina, Educational Testing Service (ETS) validation procedures and the Uniform Guidelines will be briefly reviewed.

South Carolina Decision

On January 16, 1978, the United States Supreme Court announced it had summarily affirmed the April, 1977 decision of the Federal District Court upholding South Carolina's use of the NTE in the State's certification system.

Five justices affirmed the decision of the Federal District Court of South Carolina concerning use of the NTE for certification. They did not, however, issue an opinion for their reasoning. Justices White and Brennan did, however, record a strong dissent.

Furthermore, the case has not been cited as authority in any other Federal or State court opinion with regards to the use of NTE for certification purposes.

Some of the reasons the South Carolina Court used to enable the State to use the NTE were:

1. There was no intent to discriminate against blacks in developing the testing system and that rather than a racial classification, the system created a permissable classification based on knowlege or the lack of knowledge.

2. The State's interest in hiring at least minimally competent teachers was related in a rational manner to its use of the NTE scores.

3. The use of NTE scores was a valid "business necessity." This means that in some instances a discriminatory practice might be tolerated if it were essential to the operation of the business. The "business necessity" rationale used by the South Carolina Court was based on a Supreme Court decision which said:

> The Supreme Court intended on examination of the alternatives available with respect to the legitimate employment objective identified by the employer to determine whether there is available to the employer an alternative practice that would achieve his business purposes equally well but with a lesser disparate impact by race. In examining alternatives, the risk and cost to the employer are relevant. (Griggs v. Duke Power Co.)

In the United States v. South Carolina case, the Plaintiffs presented only one alternative, graduation from an approved teacher training program, to the use of the NTE for certification purposes. The trial Court did not feel that the alternative would achieve the State's purpose in certifying minimally competent teachers as well as the use of the NTE. The Court in support of this finding made two points. One, evidence demonstrated that the teacher

training programs varied in admission requirements, academic standards, and grading practices. Two, evidence demonstrated that the State approves only general subject matter areas covered by the programs, not the actual course content of the programs. Both of these points would seem to weigh negatively on the Court's position that validation against the teacher training programs was sufficiently reflective of actual knowledge needed for the teaching positions. Here the Court would seem to be admitting that the twenty-five teacher training programs were in fact different and therefore not all would be to the same degree reflective of knowledge needed to competently perform the job. The Court, however, while finding the teaching programs themselves an inadequate measure of teacher competency saw no inconsistency in finding test validation against those same teacher programs acceptable.

ETS Validation Procedures

ETS (1976) in their Brief Amicus Curiae filed in the District Court of South Carolina stated:

> Any general challenge by the plaintiffs to the use of the NTE by South Carolina for measuring the "academic preparation" of prospective teachers would be foreclosed by the North Carolina decision. There the three-judge court held that:
>
> > beyond argument . . . the State of North Carolina has the right to adopt academic requirements and written achievement tests designed and validated to disclose the minimum amount of knowledge necessary to effective teaching. 400 F. Supp. at 348. (p. 4)

ETS (1976) describes the validation process as follows:

> The Content Review Panels were asked to make four independent types of judgments about the congruence of

> the NTE tests they were reviewing and the curriculum in teacher training programs in South Carolina. Judges were asked to indicate, on the basis of a review of the major content topics of the test, whether the emphasis in the curriculum on each content topic was about the same as, more than, or less than the percentage of the test devoted to the content topic. As an additional means of assessing the overall similarity between tests and curricula, judges were also asked to indicate one of four characterizations of greater or lesser similarity between the test and the curriculum of their institution. As a measure of the comprehensiveness of the tests, judges were asked to indicate which, if any, major content topics in the teacher education sequence were not included among the test content topics. Finally, the judges were asked to review each question on the test to determine whether the individual question was appropriate for use in South Carolina. The questions on the NTE were regarded as "content appropriate" if at least 51% of the judges indicated that at least 90% of the students taking the test would have had an opportunity to learn the answer as part of their teacher education program or in courses prerequisite to entering it. (pp. 17-18)

This content validation process was upheld by courts in North Carolina and South Carolina when they ruled in favor of using the NTE for certification in these States.

Uniform Guidelines

An important point concerning the South Carolina decision is that the decision was made before the Uniform Guidelines on Employment Selection Procedures were adopted in 1978 by the Equal Employment Opportunity Commission, the Civil Service Commission, the Department of Labor, the Department of Justice and Treasury. These Guidelines are intended to assist private employees, government contractors, employment agencies, labor organizations, state and local governments, and the Federal Government in complying with Federal law prohibiting discrimination by race, color, religion, sex, and national origin. Employment decisions such as hiring,

promotion, demotion, membership, referral, retention, licensing and certification are encompassed within the scope of the Guidelines.

The Uniform Guidelines do not require validation studies where a selection method does not adversely affect (An adverse effect is a selection rate for any racial, ethnic, or sex group which is less than 80% of the rate for the highest scoring group.) a protected group. Where validation is required, however, content, construct and criterion-related validity studies are equally acceptable.

Section 5, Uniform Guidelines on Employee Selection Procedures, General Standards for Validity Studies states:

> Evidence of the validity of a test or other selection procedure by a criterion related validity study should consist of empirical data demonstrating that the selection procedure is predictive of or significantly correlated with important elements of job performance. Evidence of the validity of a test or other selection procedure by a content validity study should consist of data showing that the selection procedure is a representative sample of important work behavior to be performed on the job for which the candidates are to be evaluated. Evidence of the validity of a test or other selection procedure through a construct validity study should consist of data showing that the procedure measures the degree to which candidates have identifiable characteristics which have been determined to be important is successful performance in the job for which the candidates are to be evaluated.

The underlining was done by this writer to indicate that regardless of the type of validity study one uses, the test one uses for certification has to be directly job related. Furthermore, Section 14, Uniform Guidelines on Employee Selection Procedures, Technical Standards for Content Validity states:

> Where a training program is used as a selection procedure and the content of a training program is justified on the basis of content validity, it should

be justified on the relationship between the content of the training program and the content of the job.

In reference to the above guidelines section, Paul Tractenberg (1980), Professor of Law, Rutgers University states:

> The South Carolina NTE case does seem to suggest that content validity may be sufficient and does seem to suggest even that content validity may be sufficient if it is related to training programs; that is, the test items connect up to the training programs rather than to the job on the other end. That may be the only court decision that says really quite that, and as such, it is both a reasurrance and a matter of some concern that it will become the uniform rule of the courts, especially in light of the new Uniform Guidelines on Employee Selection Procedures which have a provision that seems to link content validity to what happens on the job, not simply what happened in the training program.

Legal Rationale for Arkansas Procedures

A paramount issue, in addition to measurement concerns, in the Arkansas NTE study was to make the procedures and the final results as legally defensible as possible.

The project director, after consulting with several attorneys, believes that the ETS validation procedures used in the other "NTE States" would not be upheld in a current court case. The primary reasons for this belief are the Uniform Guidelines on Employee Selection Procedures and court decisions made after the South Carolina NTE case.

In Ensley v. Branch of N.A.A.C.P. v. Seibels, 616 F. 2d 812 (5th Cir. 1980) there is a broad language indicating that validation to a training program is insufficient. The court says:

> We do not believe the Davis rationale can be extended . . . to the general proposition that any test can be validated by showing a relationship to training.

> We decline to extend the Davis rationale by holding that any test can be validated against training, without respect to the test's ability to predict job performance. Such an extension would violate the requirement of job performance validation enunciation in Griggs and Abermarle as well as agency guidelines (Uniform Guidelines) elaborating upon that requirement.

A case which explicitly says that the Equal Employment Opportunity Commission (EEOC) guidelines should apply is United States v. Virginia, 620 F.2d 1018 (4th Cir. 1980). In Virginia the state was giving a test for employment of state police officers. The state had validated the test according to the content of the training programs for the job. The court said:

> In Davis the court held that a positive relationship between the test and training-course performance was a sufficient validation under standards similar to those of Title VII. While the evidence presented to the district court was sufficient to support its conclusion under Davis, upon demand the court should reconsider this issue in the light of the applicable EEOC guidelines to determine whether a contrary conclusion is required.

The reader should realize that the Davis case was used by the South Carolina Court in their NTE decision. Furthermore, the Davis case was also rendered before the Uniform Guidelines were published.

The Arkansas NTE validation procedure was designed to meet the intent of the general validity standards of the Uniform Guidelines. In other words, the validity question was designed to be job related, instead of training program related.

Any state using the NTE may, however have legal problems, regardless of the validation procedures. A recent test validation case is Guardians Association of the New York City Police Department, Inc. v. Civil Service Commission, 630 F. 2d 79 (1980). The court

in this case stated five steps for test validation. They were:

1. The test-makers must have conducted a suitable job analysis;

2. they must have used reasonable competence in constructing the test itself;

3. the content of the test must be related to the content of the job;

4. the content of the test must be representative of the content of the job; and

5. a scoring system must be used which usefully selects from among the applicants those who can better perform the job.

Directions to the Judges

The actual directions used by the judges in the Arkansas study were:

> The first rating you will make concerns item Relevance. This will be used for test validation. In order to make this judgment, you should read the item, the "correct" answer, and the distractors. (The correct answer is underlined in the test booklet.) You should then judge the relevance of the content measured by the question with respect to the domain of knowledge you believe a minimally qualified entry-level person in the certification area <u>should possess</u>.
>
> If you believe the content of the question is irrelevant to the domain of knowledge a minimally qualified entry-level person in this field should possess, then you should fill-in circle 1 on your answer sheet in the Relevance column to signify "Not Relevant."
>
> If you believe the content of the question is of doubtful or questionable relevance to the domain of knowledge a minimally qualified entry-level person in this field should possess, then you should fill-in circle 2 on your answer sheet in the Relevance column to signify "Questionable."
>
> If you believe the content of the question is important, but not quite crucial, to the domain of knowledge a minimally qualified entry-level person in this field should possess, then you should fill-in circle 3 on your answer sheet in the Relevance column to signify "Important."

If you believe the content of the question is of crucial importance to the domain of knowledge a minimally qualified entry-level person in this field should possess, then you should fill-in circle 4 on your answer sheet in the Relevance column to signify "Crucial.

Points to consider for the relevance ratings:

A. A minimally qualified entry-level person will have completed the training requirements as specified by Arkansas Certification Requirements. The requirements were sent to you for your NTE area.

B. A minimally qualified entry-level person has the minimal amount of knowledge to initially perform in the certification area.

C. In making your relevance judgment, be sure to think in terms of the full range of grade levels and/or subject areas and/or types of schools for your certification. Examples:

   a. If you are judging the NTE Elementary Education for Elementary (1-6) Certification and you find an item "Not Relevant" for early elementary grade teachers, but "Crucial" for later elementary grade teachers, then you would mark the item "Crucial."

   b. If you are judging the NTE Math examination for Math Certification and you find an item "Questionable" for teaching 9th or 10th grade math, but "Important" for teaching 11th or 12th grade math, then you would mark the item "Important."

   c. If you are judging for the Elementary Principal Certification and find an item "Crucial" for a large urban elementary school, but "Questionable" for a small town elementary school, then you should mark the item "Crucial."

   d. If you are judging the NTE in Biology and General Science, then you should respond for Biology <u>or</u> General Science Certification. If you are judging the NTE in Chemistry and Physics, then you should respond for Chemistry <u>or</u> Physics <u>or</u> Physical Science Certification.

The second judgment you will make about each item will help determine the cut-score. You should imagine a hypothetical person, who in your judgment, has the minimum amount of academic knowledge to complete the preparation program required for certification in Arkansas and has the minimum amount of knowledge to perform in the field designated by the NTE area test. With this hypothetical person in mind, you are to estimate the probability that this minimally competent person would know the answer to the NTE item without guessing. Another way of thinking about this estimation process is to think of a group of minimally competent persons and then estimate the percent of minimally competent persons who would answer the NTE item correctly without guessing.

Before you make your estimate about the item, you should also realize the item difficulty based on the NTE norm group for the item. The item difficulty or the percent who have passed the item is written beside the item in the booklet.

You should mark your estimate for each item on the response sheet under the Probability column. You should use the following scale for these estimates:

Fill in circle 1, if your estimate is between .00 - .10
Fill in circle 2, if your estimate is between .11 - .20
Fill in circle 3, if your estimate is between .21 - .30
Fill in circle 4, if your estimate is between .31 - .40
Fill in circle 5, if your estimate is between .41 - .50
Fill in circle 6, if your estimate is between .51 - .60
Fill in circle 7, if your estimate is between .61 - .70
Fill in circle 8, if your estimate is between .71 - .80
Fill in circle 9, if your estimate is between .81 - .90
Fill in circle 10, if your estimate is between .91 - 1.00

Data Analysis

This section of the paper presents the methods used to determine item validity and cut-scores. The attached chart presents the actual results from the Arkansas study.

Validity

The validity of each item was determined by computing an item mean for each item on the relevance scale. This scale had a range from one (Not Relevant) to four (Crucial). In order for an item

to be considered valid, the mean score on the relevance scale had to be greater than 2.5. In other words, the item had to be rated by the judges as closer to the important category than to the questionable category. If half of the judges had rated the item questionable and the other half had rated the item important, then the item would not have met the validity criterion since the mean rating would have been 2.50.

Cut-Scores

The cut-scores for each area examination was determined by a slight modification of a procedure known as the Angoff method. The first step for determining the cut-score was to determine an item mean on the probability scale. Since the judges had responded to a probability range for each item, the mid-point of the range was used to compute the item mean. For example, a one on the probability scale was converted to .05 since one represented the probability between .0 through .1.

The raw score cut-score for each area examination was computed by summing the mean probabilities for only the items that had met the validity criterion. A conversion formula was then used to convert the raw scores to NTE standard scores.

Chart 1

NTE Area	Number Not Valid	Number Valid	Derived Cut-Score	Percentiles	Arkansas Certification Cut-Score	Percentiles
Art Education	55	95	500	19	450	8
Biology and General Science	19	141	580	29	540	17
Business Education	35	125	590	44	550	28
Chemistry, Physics, and General Science	29	121	560	41	520	26
Early Childhood Education	28	122	560	32	500	18
Education in the Elementary School	21	129	550	28	500	16
English Language and Literature	42	108	530	26	490	15
French	55	125	530	29	490	19
Guidance Counselor	8	147	610	39	560	23
Home Economics Education	26	124	570	30	520	16
Industrial Arts Education	44	106	580	31	550	23
Introduction to the Teaching of Reading	12	138	560	28	510	16
Mathematics (Mathematics)	22	98	570	46	530	30
Mathematics (Basic Math)			520	26	470	11
Media Specialist-Library and Audiovisual Service	24	126	590	19	540	10
Music Education	30	120	570	36	510	18
Physical Education	31	119	590	40	540	22
Reading Specialist	16	134	600	32	550	17
Social Studies	36	114	550	36	500	20
Spanish	44	116	520	23	470	11
Speech Pathology	32	118	630	29	570	11
Educational Administration and Supervision (Elementary School Principal)	38	112	540	19	490	9
Educational Administration and Supervision (Secondary School Principal)	44	105	560	25	510	12
Educational Administration and Supervision (Administrator)	27	123	570	27	520	15

NTE Sample Items
Education in the Elementary School

1. A teacher is most likely to be effective in encouraging children to improve their handwriting by having them do which of the following?

   (A) Work on the letters that are circled on their papers.
   (B) Keep samples of their handwriting for comparison purposes.
   (C) Practice the basic ovals and slants for the difficult letters.
   (D) Read each other's papers and identify illegible letters.
   (E) Recopy written work in their best handwriting.

2. An intermediate-grade teacher who wants to develop pupils' interest in poetry could most profitably use poems in the short, humorous style of

   (A) Edward Lear
   (B) James Whitcomb Riley
   (C) Henry Wadsworth Longfellow
   (D) Rachel Field
   (E) Rudyard Kipling

3. In which of the following ways may a teacher best provide for the development of reading skills in an upper elementary class of thirty pupils whose reading abilities range from second to eighth grade?

   (A) By grouping the children into three levels of basic readers
   (B) By arranging for those pupils who are below grade level to go into a lower grade for reading instruction
   (C) By grouping the children into six levels of basic readers
   (D) By providing instruction in the sixth-grade reader for all children
   (E) By placing the children into skill groups and using an individualized approach

4. A pupil has become proficient in independent word attack when he has learned which of the following:

   (A) The phonics principals presented in the basal reader
   (B) How to use a decoding system that is helpful
   (C) How to use context clues in analyzing unknown words
   (D) The sound-symbol relationships of the English language
   (E) How to recognize words from configuration clues

5. Engaging in creative art activity is likely to increase a child's appreciation of art because

(A) the process of creation sensitizes him to many aspects of the completed product
(B) every activity in art generates so many overtones that appreciation is inevitable
(C) reference to actual examples helps the child in his creative work
(D) the physical activity involved sharpens perception
(E) creation and appreciation are fundamentally the same

6. After dinner, 3/4 of the pie was left. Mary then ate 1/2 of what was left. How much did she eat?

Which of the following mathematics operations is used to solve the problem above?

(A) Addition of fractional numbers
(B) Division of fractional numbers
(C) Division of whole numbers
(D) Multiplication of fractional numbers
(E) Subtraction of fractional numbers

7. After completing a unit on Native Americans, an intermediate-grade teacher asks his class to indicate on a topographical map of a mythical country where various tribes might have lived. His objective probably is to

(A) help the children build theories about peoples and their housing patterns
(B) see whether the children know how to read a map
(C) see whether the children have learned where the Native Americans lived
(D) see whether the children can match modes of adaptation to environmental areas
(E) help the children develop more accurate map skills

8. A middle-grade teacher who wishes to update and improve her teaching of science is most likely to do so by

(A) placing more stress on reading in the basic science texts
(B) including more lectures to save time for experiments
(C) guiding children to discover learnings for themselves
(D) demonstrating more experiments
(E) requiring children to write up experiments in greater detail

9. Music can be effectively related to social studies for which of the following reasons?

(A) Music has a place in the cultural and nationalistic heritage of all nations.
(B) Folk music is a reliable source of historical data among nations.
(C) Music is the best source of information about a people's culture.
(D) All nations have produced working and communal songs that can be used in the schools.
(E) Nationalistic differences in music correspond to political differences among nations.

10. According to usual patterns of child growth and development, most 6-year-olds could be expected to be able to do which of the following?

(A) Jump rope rhythmically
(B) Throw, catch, bounce, and bat a ball
(C) Roller-skate and ride a bicycle
(D) Climb a rope
(E) Hop on one foot, skip, and jump

Item 1	p = .60	Rel. 1 2 3 4	Prob. .05 .15 .25 .35 .45 .55 .65 .75 .85 .95
Item 2	p = .40	Rel. 1 2 3 4	Prob. .05 .15 .25 .35 .45 .55 .65 .75 .85 .95
Item 3	p = .78	Rel. 1 2 3 4	Prob. .05 .15 .25 .35 .45 .55 .65 .75 .85 .95
Item 4	p = .55	Rel. 1 2 3 4	Prob. .05 .15 .25 .35 .45 .55 .65 .75 .85 .95
Item 5	p = .38	Rel. 1 2 3 4	Prob. .05 .15 .25 .35 .45 .55 .65 .75 .85 .95
Item 6	p = .73	Rel. 1 2 3 4	Prob. .05 .15 .25 .35 .45 .55 .65 .75 .85 .95
Item 7	p = .58	Rel. 1 2 3 4	Prob. .05 .15 .25 .35 .45 .55 .65 .75 .85 .95
Item 8	p = .62	Rel. 1 2 3 4	Prob. .05 .15 .25 .35 .45 .55 .65 .75 .85 .95
Item 9	p = .43	Rel. 1 2 3 4	Prob. .05 .15 .25 .35 .45 .55 .65 .75 .85 .95
Item 10	p = .52	Rel. 1 2 3 4	Prob. .05 .15 .25 .35 .45 .55 .65 .75 .85 .95